Sets and Subsets

T. W. Leung

It is well known that a set with n elements contains altogether 2^n subsets, (including the empty set and itself as subsets). However if we put restrictions on the choices of subsets, say each subset may contain at most k elements, or any two chosen subsets may have non-trivial (or trivial) intersection, then our choices will be more limited. Conversely if we can select from the subsets of a set with n elements a certain number of subsets, and these subsets satisfy some pre-defined properties, then n probably cannot be too small. Or perhaps we can deduce any two of these subsets must have non-trivial intersection, etc.

There have already been many results relating sets and subsets. On the other hand, as problems of this kind may test students' analytical ability, and the required knowledge in solving this kind of problems is usually minimal, they appear in mathematical contests quite frequently.

Example 1: (Soviet Union Mathematical Olympiad 1965) A committee had 40 meetings, with 10 committee members attended each meeting. Every two committee members attended a meeting together at most once. Show that there are more than 60 members in the committee.

Solution: Every meeting was attended by 10 members. Thus there were $C_2^{10} = 45$ pairs of committee members. According to the condition given each pair would not appear in another meeting, hence for 40 meetings, there were $40 \times 45 = 1800$ distinct pairs of committee members. Now the committee had n members and there were $C_2^n = \frac{n(n-1)}{2}$ distinct pairs. We must have $1800 \le \frac{n(n-1)}{2}$, solve to get $n > 60$. □

Using the set language, the committee members form a set. Members attending a particular meeting is a subset of this set. The condition implies the intersection of any two of these subsets contains at most 1 element. Now we have altogether 40 such subsets, and we can deduce that the size of the committee cannot be too small. One strategy is solving this type of problems is to look for a suitable observable quantity, then investigate this quantity using different perspectives. In this case the quantity we look for is the number of "pairs" of committee members attending a same meeting.

Alternative Solution of Example 1: The problem may be tackled as follows. Each committee was consisted of 10 members, hence altogether $40 \times 10 = 400$ "persons" attended the meetings. Suppose the committee had at most 60 members, since $\frac{400}{60} \approx 6.67$, we must have 1 member in the committee who attended at least 7 meetings. But from the condition given, the members (except this fellow) attending these 7 or more meeting must not be the same. Thus there are at least $7 \times 9 = 63$ members, a contradiction! (Note implicitly we have used the pigeon hole principle.) □

Example 2: (Austrian-Polish Mathematics Competition 1978) There are given 1978 sets, each containing 40 elements. Every two sets have exactly one element in common. Prove that all 1978 sets have a common element.

Solution: Suppose A is one of the sets. Consider the other 1977 sets, each set has a common element with A. Since $\frac{1977}{40} \approx 49.43$, there exists $x \in A$ which also belongs to 50 other sets $A_1, A_2, \cdots, A_{50}$ say, and x is the only common element of these sets. Consider now another set B. If $x \notin B$, as B intersects non-trivially with $A_1, A_2, \cdots, A_{50}$ and the intersecting elements cannot be common, hence B contains at

least 50 elements, a contradiction. Thus we must have $x \in B$. B is arbitrary, hence x is in every other subset. The proof is complete. □

The result can be extended and the proof is entirely analogous: Suppose there are n sets, with each set containing k elements. Every two sets contain exactly one common element. If $n > k^2 - k + 1$, then the n sets contain a common element.

We now consider a more difficult example.

Example: (Soviet Union Mathematical Olympiad 1996) 16000 committees were formed with members coming from 1600 councillors. Each committee consisted of exactly 80 members. Show that there exist two committees, and they had at least 4 common members.

Solution: In this problem we do not consider pairs of members in each committee. Instead we consider from the perspective an individual councillor, and see how many pairs of committees he joined. Suppose councillors 1, 2, …, 1600 joined respectively $k_1, k_2, \cdots, k_{1600}$ committees. Hence altogether there are $C_2^{k_1} + C_2^{k_2} + ... + C_2^{k_{1600}}$ "committee pairs". Now there are N committees, ($N = 16000$), then $k_1 + k_2 + \cdots + k_{1600} = 80N$. We estimate these committee pairs

$$C_2^{k_1} + C_2^{k_2} + ... + C_2^{k_{1600}} = \frac{k_1(k_1-1)}{2} + \frac{k_2(k_2-1)}{2} + ... + \frac{k_{1600}(k_{1600}-1)}{2}$$

$$= \frac{k_1^2 + k_2^2 + ... + k_{1600}^2}{2} - \frac{k_1 + k_2 + ... + k_{1600}}{2} \geq \frac{1}{2}\frac{(k_1 + k_2 + ... + k_{1600})^2}{1600} - \frac{k_1 + k_2 + ... + k_{1600}}{2}$$

$$= \frac{1}{2}\frac{(80N)^2}{1600} - \frac{80N}{2} = 2N^2 - 40N = 2N(N-20)$$

Suppose now every two committee have at most three common members, then can only be at most $3C_2^N = \frac{3N(N-1)}{2}$ committee pairs. Hence $2N(N-20) \leq \frac{3}{2}N(N-1)$, or $N \leq 77$, contradicting $N = 16000$. □

(Note that in the estimate we have made use of the Cauchy-Schwarz Inequality.)

Curiously enough, we found a similar example.

Example 4: (International Mathematical Olympiad 1998) In a contest there are m candidates and n judges, where $n \geq 3$ is an odd integer. Each candidate is evaluated by each judge as either pass or fail. Suppose that each pair of judges agrees on at most k candidates. Prove that $\frac{k}{m} \geq \frac{n-1}{2n}$.

Solution: The problem has reminded us to consider pairs of judges who agree on particular candidates. Consider candidate i, $1 \leq i \leq m$, suppose x_i judges consider him pass and y_i judges think he fail. Then the number of pairs of judges that agree on this candidate is

$$C_2^{x_i} + C_2^{y_i} = \frac{x_i(x_i-1)}{2} + \frac{y_i(y_i-1)}{2} = \frac{x_i^2 + y_i^2}{2} - \frac{x_i + y_i}{2} \geq \frac{1}{2}\frac{(x_i + y_i)^2}{2} - \frac{x_i + y_i}{2}$$

$$= \frac{1}{4}n^2 - \frac{n}{2} = \frac{1}{4}\left[(n-1)^2 - 1\right].$$

Now n is odd and $C_2^{x_i} + C_2^{y_i}$ is an integer, hence $C_2^{x_i} + C_2^{y_i}$ is at least $\frac{1}{4}(n-1)^2$. Also there are n judges and any pair of judges agree on at most k candidates, there are at most kC_2^n pairs that agree on certain candidates. Thus $kC_2^n \geq \sum_{i=1}^{m}\left[C_2^{x_i} + C_2^{y_i}\right] \geq \frac{m(n-1)^2}{4}$, simplify to obtain the desired result. □

We now consider a problem of slightly different flavor. Consider sequences of length n and such that each sequence contains only 0s or 1s. Define the "distance" between two such sequences as the number of positions they differ. For example 1101011 and 1011000 are two sequences of length 7, and they differ on positions 2, 3, 6 and 7, hence the distance between them is 4. Using set language, we say that a set contains 7 elements 1, 2, 3, 4, 5, 6 and 7, as sequence 1 are non-zero at positions 1, 2, 4, 6, and 7, and we may consider sequence 1 as a subset containing elements 1, 2, 4, 6 and 7. Likewise sequence 2 is a subset containing 1, 3 and 4. Elements belong to sequence 1 or sequence 2, but not both, in this case consist of 2, 3, 6 and 7, together form a subset, called the "symmetric difference" of the two subsets. One can verifies easily that the distance between the sequences corresponds precisely to the number of elements in the symmetric difference. Now our concern is, given n and a restriction on the distances between the sequences, at most how many sequences we can get?

Example 5: Given m sequences of length n and the sequences contain only 0s and 1s. Suppose the distance between any two such sequences is at least d, then $m \leq \frac{2d}{2d-n}$.

Solution: Consider any two sequences and the corresponding pairs that differ. First we note that there are C_2^m pairs of sequences. For any pair of sequences, the number of pairs that differ or "distance" between them is at least d. Hence the sum of total distances between all possible pairs of sequences is at least dC_2^m. Now we list all of these m sequences as m rows. Then each column j, $1 \leq j \leq n$, correspond to the j^{th} position of these m sequence. If the j^{th} column contains x_j 0s, then it will contain $m - x_j$ 1s. This column will contribute $x_j(m-x_j)$ pairs of positions that the sequences differ. Observe

$$x_j\left(m-x_j\right) \leq \left[\frac{x_j + \left(m-x_j\right)}{2}\right]^2 = \frac{m^2}{4},$$

thus $dC_2^m \leq \sum_{j=1}^{n} x_j\left(m-x_j\right) \leq \sum_{j=1}^{n} \frac{m^2}{4} = \frac{nm^2}{4}$. Simplify to get $m \leq \frac{2d}{2d-n}$. □

For instance, if $n = 7,\ d = 4,$ then $\frac{2d}{2d-n} = 8.$ It is impossible to construct 9 sequences if the minimum distance between any pair is 4. (Our readers perhaps would like to construct 8 such sequences.) In fact this result is a special case of a theorem in Coding Theory, usually referred as a Plotkin Bound.

There are many interesting results concerning sets and subsets for us to investigate.